Bearing Fault Diagnosis under Varying Work Conditions Based on Synchrosqueezing Transform, Random Projection, and Convolutional Neural Networks

Boubker Najdi[1], Mohammed Benbrahim[2], Mohammed Nabil Kabbaj[3]

[1,2,3] *Systems Engineering, Modeling and Analysis Laboratory, Faculty of Sciences, University Sidi Mohamed Ben Abdellah, Fez, Morocco.*
aboubakr.najdi@usmba.ac.ma
mohammed.benbrahim@usmba.ac.ma
n.kabbaj@usmba.ac.ma

ABSTRACT

Bearings are critical components in rotating machinery, and their failure can lead to costly repairs and downtime. To prevent such failures, it is important to detect and diagnose bearing faults early. In recent years, deep learning techniques have shown promise for detecting and diagnosing bearing faults automatically. While these algorithms can all achieve diagnostic accuracy of over 90%, their generalizability and robustness in complex, extreme variable loading conditions have not been thoroughly validated. In this paper, a feature extraction method based on Synchro-squeezing Wavelet Transform (SSWT), Random projection (RP), and deep learning (DL) is presented. To fulfill the data requirements of neural networks, data augmentation is initially utilized to augment the size of the original data. Subsequently, the SSWT technique is employed to convert the signals from the Time domain to the Time-Frequency domain, resulting in the conversion of the 1-D signal to a 2-D feature image. To decrease the complexity of deep learning computation, data preprocessing involves utilizing Random projection to reduce feature dimensionality. The final step involves constructing a Convolutional Neural Network (CNN) model that can identify fault features from the obtained Time-Frequency images and perform accurate fault classification. By utilizing the CWRU and IMS datasets to evaluate the method, the study demonstrates that the suggested approach outperforms advanced techniques in terms of both diagnostic accuracy and robustness.

1. INTRODUCTION

The rapid progress in science and technology has significantly advanced modern industry (Gai, Shen, Hu, & Wang, 2020). This advancement has led to the creation of highly integrated, precise, and intelligent rotating machinery, essential in various sectors of modern manufacturing. Bearings are crucial components in these machines, providing stability and enabling the shaft to rotate relative to a stationary part. Over time, the heavy loads exerted on these components lead to their degradation, potentially causing unexpected equipment failures.

According to studies (Gai et al., 2020), bearing issues are responsible for about 40% of motor breakdowns, leading to substantial financial losses, increased maintenance costs, extended downtime, and even risks to human safety (Liang, Zuo, & Feng, 2018). Therefore, effective rolling bearing fault diagnosis and condition monitoring are critical.

Many academics are integrating various signal processing techniques with machine learning algorithms to address challenging research problems, driven by the industry's eagerness to adopt advanced digital technologies. The most common meth-od for fault classification and diagnosis in rolling bearings is vibration feature extraction. Initial research efforts focused on calculating different statistical indices in the time domain (Renwick & Babson, 1985; Altman & Mathew, 2001) and frequency domain (Hu, 2006; Lei, 2008). However, variations in operating speed can render the bearing environment unstable, making stationary signal analysis methods less effective. This challenge has led to the increased use of Time-Frequency Analysis (TFA) approaches, which provide both time and frequency information, making them suitable for non-stationary signals (Ricci & Pennacchi, 2011). Classical TFA techniques include linear and quadratic approaches, such as the Short-Time Fourier Transform (STFT),

Wavelet Transform (WT), and Wigner-Ville Distribution (W-VD). The Time-Frequency Representation (TFR) produced by conventional linear TFA methods, however, often experiences spreading around the instantaneous frequency (IF) trajectory of the signal, a limitation imposed by the Heisenberg uncertainty principle. To address this, the reassignment method (RM) (Auger & Flandrin, 1995), which employs post-processing techniques, has been suggested. However, this method does not allow for signal reconstruction (Auger et al., 2013). Daubechies proposed the SSWT as a solution to this problem. The primary aim of the SSWT is to enhance the readability of the TFR provided by the Continuous Wavelet Transform (CWT) (Daubechies, 2000) through a reassignment technique while maintaining the invertibility of the transform. Consequently, the SSWT is increasingly used for fault diagnosis in various applications (Liu, Li, & Chen, 2018; Lee, Chen, Wu, & Jamerson, 1989; J. Wen et al., 2015).

Deep learning has revolutionized the training of neural networks with its capability to handle long and complex layers. It is widely applied in various fields, including image generation and processing (Bengio, 2016; LeCun, Bengio, & Hinton, 2015). The CNN, a deep learning model, mimics the mammalian visual system and is known for its three key architectural elements: local receptive fields, shared weights, and spatial domain pooling, making it particularly effective for two-dimensional visual input recognition (Pham, Kim, & Kim, 2021b; Zan, Wang, Wang, Liu, & Gao, 2019; Pham, Kim, & Kim, 2021a).

CNNs have gained popularity in bearing fault diagnosis. For instance, Wen et al. (L. Wen, Li, Gao, & Zhang, 2018) developed 12 CNN models based on the standard CNN and LeNet-5, transforming one-dimensional time series data into two-dimensional image signals for input into the LeNet-5, demonstrating success in three mechanical fault classification tasks. Wang et al. (Wang, Zhuang, Duan, & Cheng, 2016) enhanced CNNs' generalization for fault detection by applying Morlet wavelet decomposition, bilinear interpolation, and rectified linear units to grayscale images derived from vibration signals. Another approach, combining Cyclic Spectral Coherence (CSCoh) with CNNs, has been proposed for diagnosing rolling bearing faults, showing improved detection performance (Chen, Mauricio, Li, & Gryllias, 2020). However, these methods pose challenges in real-world engineering due to the high demand for training samples. Gathering fault data is particularly difficult since equipment typically operates without failure, leading to a data imbalance in fault identification.

To address the computational and storage burden of high-dimensional features in multi-layer CNNs, the Random Projection (RP) method is employed for image pre-processing and dimensionality reduction. RP, simpler and faster than traditional techniques like PCA, becomes increasingly advantageous as dimensions grow (Bingham & Mannila, 2001), (M.Benbrahim, 2014).

Considering the literature review in the introduction and to address the aforementioned limitations, this research integrates several techniques, including SSWT, data augmentation, RP, and CNN, to develop a comprehensive deep neural network framework for diagnosing bearing faults. In our proposed model, SSWT is employed to enhance the readability of the TFR. RP effectively reduces the computational overhead associated with multi-layer CNN features and removes redundant information from the SSWT outputs, accelerating the deep learning training process. Our model utilizes a CNN architecture inspired by the strengths of the LeNet-5, capturing hierarchical information. Data augmentation techniques are implemented to expose the model to a wider range of scenarios, thereby improving its generalization ability. This is particularly crucial in tackling the challenges associated with limited datasets in real-world engineering, where collecting faulty data is often difficult. The augmentation introduces diversity, enhancing the model's adaptability to various conditions and ultimately improving its performance in bearing fault diagnosis. This comprehensive approach leverages the strengths and mitigates the weaknesses of each technique, ensuring the effectiveness of our model.

The rest of this paper is organized as follows: Section 2 explains the background theory behind the methods used in this research. The proposed diagnostic method is presented in Section 3. Section 4 introduces the experimental process and analyzes the results. Finally, Section 5 draws conclusions.

2. THEORETICAL BACKGROUND

2.1. Convolutional Neural Networks

In this section, the theory behind CNNs is delved into, highlighting their increasing use in addressing various challenges in computer vision. CNNs are particularly adept at detecting patterns in two-dimensional feature spaces, such as images. These patterns can often be represented by templates, which are essentially smaller segments of the image itself. A standard CNN comprises three types of layers: a convolutional layer, a pooling layer, and a fully connected layer (FCL), similar to those found in typical neural networks (Ghosh, Sufian, Sultana, Chakrabarti, & De, 2020), as illustrated in Figure 1a.

The convolutional layer employs a set of filters to generate feature maps from input images. These feature maps capture essential patterns and characteristics of the input. The pooling layer serves as a down-sampling mechanism, effectively reducing the dimensionality of the feature maps and retaining only the most salient features. After several alternating convolutional and pooling layers, the fully connected layers are utilized to calculate class scores, leading to the final output. In the convolutional layers, feature maps from the previous

layer are convolved with learnable kernels and then passed through an activation function to produce the output feature map. This process allows the combination of convolutions from different input maps into a single output map, further enhancing the network's ability to extract and learn complex patterns. Each output map in a CNN can include convolutions from multiple input maps, which can be computed as follows:

$$x_j^i = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right), \qquad (1)$$

In this equation, x_i represents the i-th input map, k_{ij}^l is a convolutional kernel for the j-th feature map at the l-th layer, $*$ denotes the convolutional operation, b_j^l is an additive bias, and M_j is the set of input maps contributing to the j-th output map at layer l. The results of this convolution are then passed through an activation function, f (Bouvrie, 2006). One of the most widely used activation functions in Neural Networks (NNs) today is the Rectified Linear Unit (ReLU), defined as:

$$ReLU(x) = \max(0, x), \qquad (2)$$

The ReLU function introduces non-linearity in the model, allowing it to learn complex patterns in the data.

In a CNN architecture, the pooling layer is typically positioned between convolutional layers. This layer plays a crucial role in reducing the number of parameters and computations in the network, thereby mitigating the risk of overfitting. It achieves this by gradually reducing the spatial size of the network. One of the most common types of pooling layers is the max-pooling layer, which selects the maximum value from each region of the input feature map. This process is achieved using filters that slide through the input, as depicted in Figure 1b, which shows the max-pooling transformation.

In the final phase of a CNN, the output feature maps from the last convolutional or pooling layer are typically flattened into a one-dimensional (1D) vector. This flattening process is essential to transform the multi-dimensional output into a format suitable for the fully connected layers that follow. In these layers, each input is connected to each output with a learnable weight. The expression for the output in a fully connected layer is given by:

$$y = f\left(\sum_{i=1}^{n} x_i w_i + b_i\right), \qquad (3)$$

where y represents the output value, x_i is the i-th input value in the fully connected layer, w_i and b_i are the weight and bias associated with x_i, respectively, and f denotes the activation function. This structure allows the network to learn complex relationships between the high-level features extracted by the
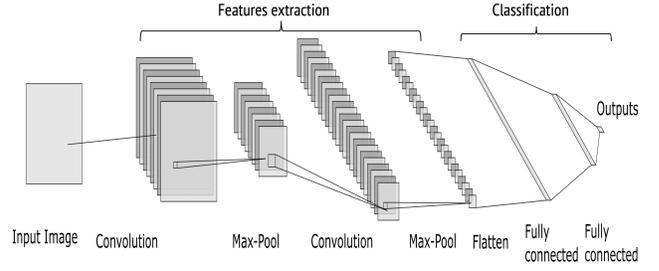
(a) A typical CNN architecture
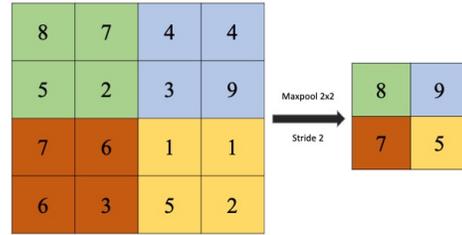
(b) Maxpooling Transformation

Figure 1. Illustration of CNN components: (a) A typical CNN architecture, (b) Maxpooling transformation

previous layers and the final output.

2.2. Random Projection

The main objective of any data transformation or projection approach is to maintain as much information as possible between the initial and transformed datasets while enhancing the data's presentation in its new form.

Even though the notion of dimensionality reduction is popular and beneficial, it has several drawbacks. First, the projection directions that arise are data-reliant, which causes issues when the size of the data grows. Second, they can be computationally too expensive. It has been demonstrated experimentally that results achieved with the RP approach are equivalent to those obtained with PCA, and that the RP method takes a fraction of the time needed by PCA (Kononenko & Kukar, 2007).

In RP the original m-dimensional data $X \in R^m$ is transformed to a n-dimensional space ($n \ll m$), using a random generated $n * m$ matrix $S = R.X$ via $S \in R$ with ($n \ll m$) whose elements are scaled to have a unit length and chosen to be independent and identically distributed (Kononenko & Kukar, 2007). The RP projects the original m-dimensional data to an n-dimensional space ($n \ll m$) while preserving the distances between the observations through the Johnson–Lindenstrauss lemma below (Johnson & Lindenstrauss, 1984):

Lemma 1 (Johnson–Lindenstrauss lemma). *For any $0 < \epsilon < 1$ and any finite set X of points in $\mathbb{R}^m$, there exists a linear map $M : \mathbb{R}^m \to \mathbb{R}^k$ into a space of dimension $k = O(\frac{\log(|X|)}{\epsilon^2})$ such that for all $u, v \in X$:*

$$(1-\epsilon)\|u-v\|^2 \leq \|M(u)-M(v)\|^2 \leq (1+\epsilon)\|u-v\|^2 \quad (4)$$

This theorem establishes a mechanism to determine the minimum dimension in which the original data can be embedded without significantly affecting the dataset's local properties. This embedding is achieved using a randomly generated matrix $M_{K \times m}$, where the elements are scaled to unit length and chosen to be independent and identically distributed. This approach ensures that, in expectation, the Euclidean distance between any two points is approximately preserved. The transformation of the original data is represented as:

$$X_{K \times N}^{RP} = M_{K \times m} \cdot X_{m \times N} \quad (5)$$

where $X_{K \times N}^{RP}$ represents the k-dimensional projected data, $M_{K \times m}$ is the random transformation matrix, and $X_{m \times N}$ is the original set of N-dimensional observations.

For the distribution of matrix M, there are several options. Achlioptas (Achlioptas, 2003) proposed an efficient method for applications such as databases. The elements m_{ij} of the matrix M can be generated according to one of the following distributions:

$$m_{ij} = \begin{cases} +1 & \text{with probability } 0.5, \\ -1 & \text{with probability } 0.5. \end{cases} \quad (6)$$

$$m_{ij} = \sqrt{3} \times \begin{cases} +1 & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -1 & \text{with probability } \frac{1}{6}. \end{cases} \quad (7)$$

In practice, any zero-mean, unit-variance distribution for m_{ij} will yield a mapping that satisfies the Johnson-Lindenstrauss lemma.

2.3. Synchrosqueezing wavelet tranform

The wavelet synchrosqueezed transform is a time-frequency technique that reassigns the energy of a signal in frequency. This reassignment serves to counteract the spreading effects induced by the mother wavelet. Notably, unlike other time-frequency reassignment techniques, synchrosqueezing precisely redistributes energy along the frequency axis while preserving the signal's time resolution.

Let ψ be a complex mother wavelet, that has strictly positive support and meets the conventional admissibility criterion $\int_0^\infty \frac{\hat{\psi}(z)}{z}\,dz < \infty$ (Daubechies, 1992) as time t and scale $a > 0$, the CWT of a signal $s(t)$ is:

$$W_\psi s(a,t) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} s(\tau)\psi^*(\frac{\tau - t}{a})\,d\tau \quad (8)$$

then the instantaneous frequencies from the CWT are extracted in equation 9, w_s, using the phase transform defined as:

$$w_s(a,t) = \frac{1}{2\pi j W_\psi s(a,t)} \frac{\partial W_\psi s(a,t)}{\partial t} \quad (9)$$

the new time-frequency representation is obtained by mapping the information from the time-scale plane to the time-frequency plane. Thus, the synchrosqueezing procedure is to move the coefficients vertically according to the map $(t, a) \longmapsto (t, w_s(a, t))$:

$$SSWTs(t, \eta) = \int_{\{(a,t):\eta, w_s(a,t)\}} \frac{1}{a^{\frac{2}{3}}} W_\psi s(a,t)\,da \quad (10)$$

where $SSWTs(t, n)$ is the SSWT of signal $s(t)$, η is the IF centered on the level of curves w_s.

To visualize how the SSWT sharpens the TF transformation and increases its resolution in the frequency domain, a basic Gaussian-modulated cosine signal is used and represented in Figure 2a, $s(t) = A \cdot \cos(2\pi f_0 t) \cdot e^{-\frac{t^2}{2\sigma^2}}$ as an example, where the amplitude of the signal $A = 1$, the frequency of the cosine term $f0 = 28Hz$, and $\sigma = 0.2$ is the standard deviation controlling the width of the Gaussian term and t is time.

Figure 2b demonstrates that the CWT does not excel in producing a concentrated time-frequency (TF) representation. In contrast, as shown in Figure 2c, the SSWT achieves superior time-frequency resolution by mitigating the spreading effect inherent in the CWT. The synchrosqueezing technique of the SSWT selectively reallocates energy, addressing the CWT's tendency to disperse signal energy, which can lead to reduced precision. This results in a more accurate and focused depiction of the signal's temporal and frequency characteristics.

2.4. Data Augmentation

To achieve a robust bearing diagnostic model, data augmentation is used on raw time-series data, implying the deliberate addition of greater diversity to the training set. By doing so, the model becomes capable of handling unexpected changes in real-world data. This deliberate diversification ensures the model's adaptability to sudden shifts, highlighting its robust nature. By strategically augmenting raw time-series data, the model's efficacy is ensured in a variety of changing scenarios while also reinforcing its overall resilience. Moreover, the overfitting of the model is prevented which increases the generalization capabilities of deep learning models by using data augmentation (Shorten & Khoshgoftaar, 2019). In this work, sample-based augmentation is used to increase the size
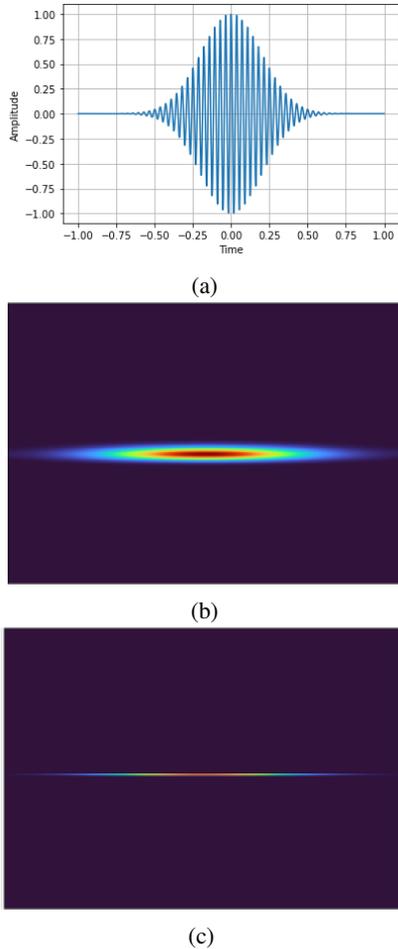
(a)

(b)

(c)

Figure 2. Comparison between the CWT (b) and SSWT (c) representations of the signal $s(t)$ (a)

of the bearing vibration dataset. The training and testing samples are expected to be taken from the original dataset, and the augmentation is done separately on each sample of the signal.

To diversify the dataset for bearing fault diagnostics and adapt the model to varying conditions, a new data augmentation approach is proposed. The flowchart of the utilized techniques is presented in Figure 3. This method applies four different augmentation techniques to the original data. Initially, the raw dataset is subjected to high levels of noise to simulate harsh operating conditions and enhance the model's resilience to noisy real-world environments. Additionally, the data undergoes three distinct augmentation methods: temporal warping, magnitude warping, and permutations, each based on applying white noise and other transformations to the original data. Time warping introduces temporal distortions, which challenge the model to identify fault patterns across various time scales, enhancing its resilience to speed variations. In contrast, magnitude warping adjusts amplitude values, aiding the model in detecting faults associated with magnitude fluctuations. Permutation disrupts the temporal

sequence of the signal's data points. This emphasizes the importance of pattern recognition independent of a specific temporal order and reduces the model's reliance on memorizing sequences, thereby improving its adaptability to unseen sequences. The augmented dataset, comprising the original dataset along with those modified by high noise, time warping, magnitude warping, and permutation, enhances the training data. This improvement increases the model's generalization capability and robustness. Figure 4 depicts the signals generated using this data augmentation approach.

3. PROPOSED METHOD FOR FAULT DIAGNOSIS

In this section, the proposed method for bearing diagnosis is presented. This method leverages a CNN architecture along with RP, enhanced with data augmentation techniques and Time-Frequency feature extraction. The overall process is illustrated in Figure 5, which provides a flowchart of the proposed bearing diagnosis method.

One of the key advantages of employing a DNN architecture in our approach is its ability to extract meaningful patterns from data through non-linear transformations and approximations. The proposed architecture is structured in two major stages:

- **Phase 1**: Four different data augmentation methods (e.g., high noise, time warping, magnitude warping, and permutation) are used to create synthetic data from the original signal. This synthetic data is then combined with the original data to form a comprehensive augmented dataset. The next step involves applying the SSWT to extract time-frequency image features. The dimensionality of the obtained images is reduced using the random projection algorithm.

- **Phase 2**: The images of the training dataset are fed to the CNN in the training phase. From these training image features, the CNN model learns to classify bearing health conditions with frequency variations over time. The testing dataset is used after the training to evaluate the model's performance.

The following subsection provides further information and theoretical background about the techniques of the suggested diagnostic method.

3.1. Case studies and results

In this section, two bearing datasets are used to verify the performance and the effectiveness of the proposed approach.

- **Case 1 (CWRU dataset)**: the Bearing Data Center at Case Western Reserve University provides the CWRU rolling bearing dataset (Smith & Randall, 2015).
- **Case 2 (IMS dataset)**: the data set was provided by the Center for Intelligent Maintenance Systems (IMS), Uni-
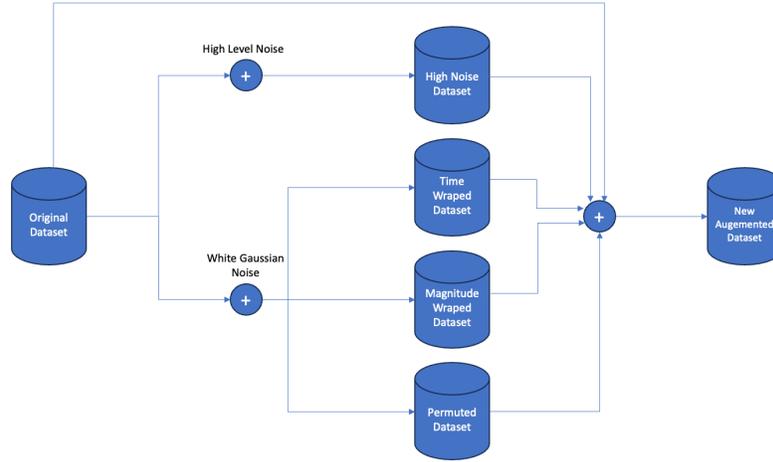
Figure 3. The flowchart of the data augmentation process.

versity of Cincinnati (J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services, 2007).

Table 1 shows the software and the hardware environment used through all the experiments:

Software	Hardware
Tensorflow 2.0 Framework	2 GPUs: Tesla V100s-PCI 32Gb
Python programming language 3.10	High performance computation server

Table 1. Software and Hardware configurations.

3.2. Case 1: CWRU dataset

The Case Western Reserve University (CWRU) dataset involved experiments using a 2 hp (approximately 1.5 Kw) Reliance electric motor. An accelerometer sensor was positioned near and far from the motor bearings to collect data. Defects in motor bearings were introduced through Electro-Discharge Machining (EDM). The setup of the CWRU test stand is depicted in Figure 6 (Smith & Randall, 2015).

This dataset includes a total of 10 health conditions, consisting of nine fault conditions and one normal condition (N). The fault conditions are Ball fault (B), Inner Race Fault (IR), and Outer Race Fault (OR), each with damage sizes of 0.007, 0.014, and 0.021 inches (where 1 inch equals 2.54 cm).

The faulty bearings of the test motor were replaced, and vibration data was recorded under motor loads ranging from 0 to 3 horsepower (motor speeds from 1797 to 1729 RPM), as detailed in Table 2. Acceleration measurements were captured over a 10-second duration with a sampling frequency of 12 KHz.

In this study, each health condition is represented by 12000

Dataset Number	Load (hp)	Speed (RPM)
1	0	1797
2	1	1772
3	2	1750
4	3	1729

Table 2. Description of the datasets under different loads/speeds.

Health Conditions	Label	Fault size(inches)	Training samples	Testing samples
N	0	–	750	180
B7	1	0.007	750	180
IR7	2	0.014	750	180
OR7	3	0.021	750	180
B14	4	0.007	750	180
IR14	5	0.014	750	180
OR14	6	0.021	750	180
B21	7	0.007	750	180
IR21	8	0.014	750	180
OR21	9	0.021	750	180
Total Samples	–	–	**7500**	**1800**

Table 3. Bearing health conditions and class labels of CWRU for each Dataset.

data points, referred to as observations. For the training dataset, 150 sample segments, each consisting of 640 points, are extracted. These samples are then augmented four times using the techniques discussed in the previous section, resulting in 750 samples for each health condition. This augmentation increases the training dataset to a total of 7500 samples. The same augmentation process is applied to the testing dataset, yielding 1800 samples. Augmenting the test set ensures that the test data closely resembles the training set, enhancing the model's performance and its ability to generalize to similar signals. Table 3 summarizes the health conditions, along with the number of training and testing samples.
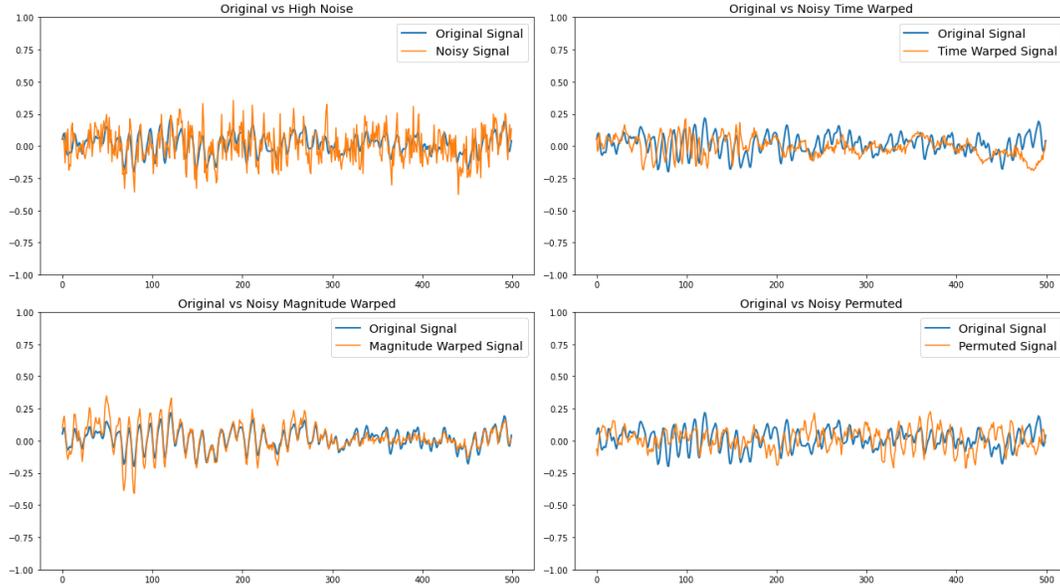
Figure 4. The original signal against the data augmentation techniques.

3.2.1. Results of signal transformation

The signal-to-image conversion results for the nine bearing faults using the SSWT Transform are depicted in Figure 7. Each converted image has a resolution of 128×640 pixels. However, this resolution is still relatively large and contains redundant information. To address this, the random projection technique previously discussed is applied to compress the images, making them more suitable for processing by the CNN model. The results of this compression are shown in Figure 8, where the randomly projected images clearly distinguish between different fault conditions. These images highlight key time and frequency features from the SSWT, providing an intuitive means to categorize the faults.

3.2.2. Analysis and Results

In the CNN architecture for bearing fault diagnosis, a multi-layered network was designed inspired by the LeNet-5 model. This network is capable of capturing hierarchical features from input data effectively. Table 4 details the specifics of our network structure. The architecture includes four convolutional layers—with 32, 64, 128, and 256 filters respectively—progressively increasing in size to capture more complex features. Following each convolutional layer, max-pooling layers reduce spatial dimensions, aiding in feature abstraction. To prevent overfitting, drop-out layers are incorporated after the first and second dense layers. Our model consists of two densely connected layers, and hyperparameter tuning was a critical step to ensure optimal performance. Grid search was conducted over predefined ranges for filters, dense units, dropout rates, and learning rates. The final model employs 256 and 1280 units for the dense layers, dropout rates of 0.3

Layer No.	Type	Filter Size	Activation	Output Shape
1	InputLayer	-	-	(128, 128, 1)
2	Conv2D	(3, 3)	ReLU	(126, 126, 32)
3	MaxPooling2D	(2, 2)	-	(63, 63, 32)
4	Conv2D	(3, 3)	ReLU	(61, 61, 64)
5	MaxPooling2D	(2, 2)	-	(30, 30, 64)
6	Conv2D	(3, 3)	ReLU	(28, 28, 128)
7	MaxPooling2D	(2, 2)	-	(14, 14, 128)
8	Conv2D	(3, 3)	ReLU	(12, 12, 256)
9	MaxPooling2D	(2, 2)	-	(6, 6, 256)
10	Flatten	-	-	9216
11	Dense	-	ReLU	1280
12	Dropout	-	-	(None)
13	Dense	-	ReLU	256
14	Dropout	-	-	(None)
15	Dense	-	Softmax	10

Table 4. The CNN architecture for bearing fault diagnosis.

and 0.1, uses the Adam optimizer (Kingma & Ba, 2014), and a learning rate of 6.8e-5. Activation functions include ReLU for convolutional and dense layers, and Softmax for classification in the output layer. The training comprises 20 epochs, split into two batches of 10 epochs each, with batch sizes of 128 and 256, respectively. This architecture strikes a balance between robustness and complexity, ensuring accurate fault diagnosis in bearing systems.

Our objective is to assess the stability and robustness of our CNN model under various conditions. Four separate datasets from the CWRU bearing dataset were used to train the CNN model, and the model's performance was monitored through training accuracies and losses. Figure 9a and Figure 9b show the training accuracy and loss curves for the 20 training epochs, with accuracy values ranging from 0% to 99%. It was ob-
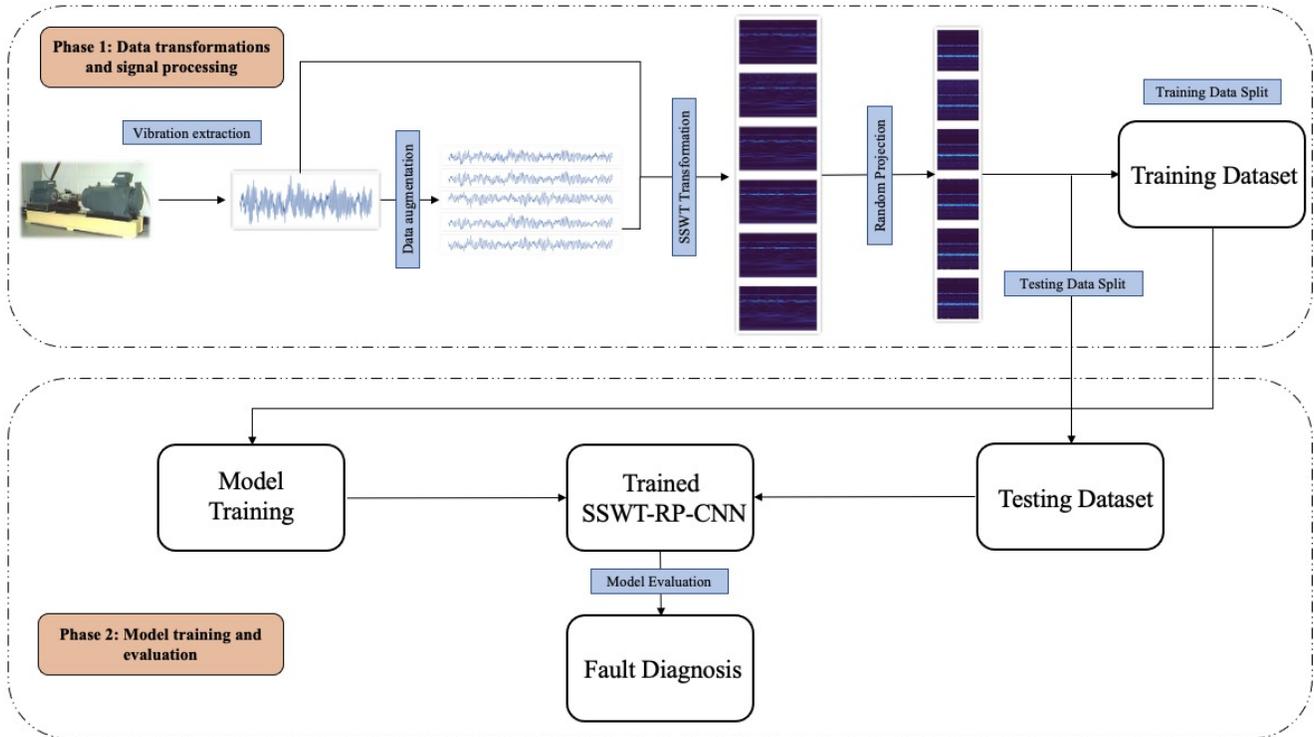
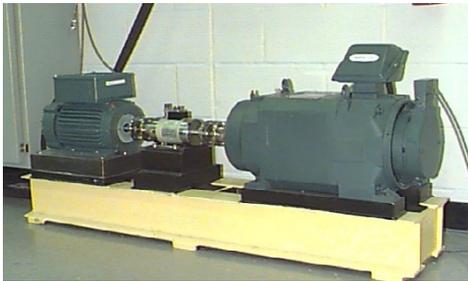Figure 5. Flowchart of the proposed bearing diagnosis method.

Figure 6. The CWRU test stand (Smith & Randall, 2015).

served that the curves began to stabilize at epoch 8, indicating consistent achievement of 100% accuracy beyond this point by the model.

The training loss curves rapidly decreased during the initial epochs, suggesting a swift learning rate. By the third epoch, the training losses had significantly reduced, implying an increase in accuracy. After epoch 8, the loss curves stabilized across all four datasets, indicating that the model had reached its optimal settings. Figure 10 demonstrates the model's perfect accuracy of 100 percent on the training set for all datasets, efficiently identifying the underlying patterns in the data.

The effectiveness of our model in capturing complex fault-related characteristics is demonstrated through t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization. t-SNE
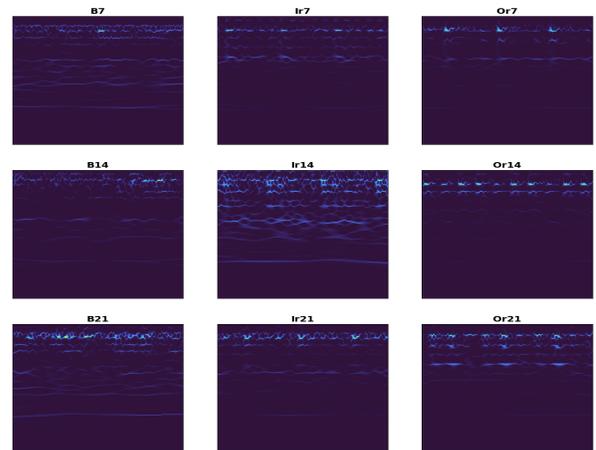

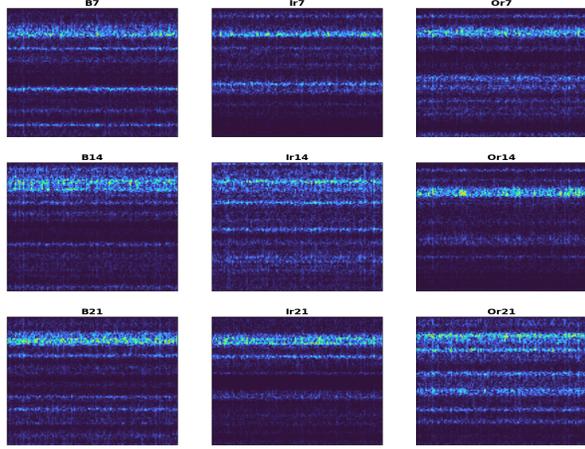Figure 7. The SSWT of all CWRU bearing faults.

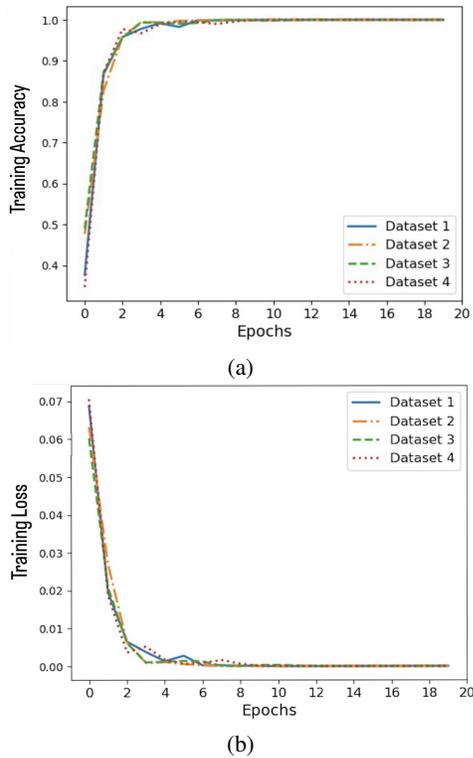Figure 8. The RP of the obtained SSWT bearing faults of CWRU.

(a)

(b)

Figure 9. Training accuracy (a) and training loss (b) curves of the four datasets(best CNN model).

is a technique used to visualize high-dimensional data in a low-dimensional space, revealing patterns not readily apparent. In the t-SNE plot of the raw data (Figure 11), no distinct clusters corresponding to different fault states are observed. The data points appear randomly distributed, suggesting a lack of clear fault-related patterns.
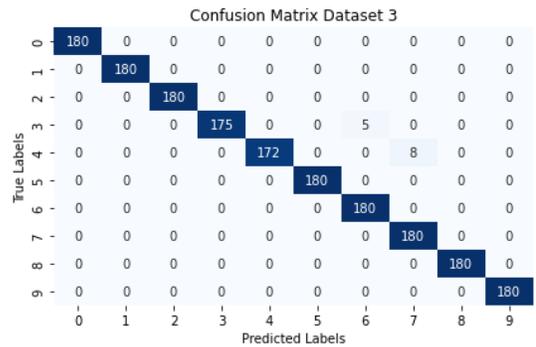
However, after applying the SSWT and RP, the t-SNE visualization begins to show distinct clusters (Figure 12). This change indicates that these preprocessing steps help in revealing fault-related structures in the data. The t-SNE visualiza-
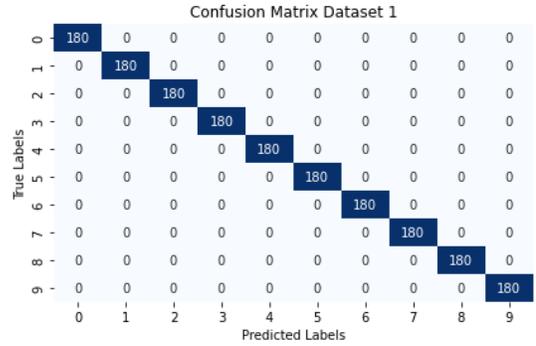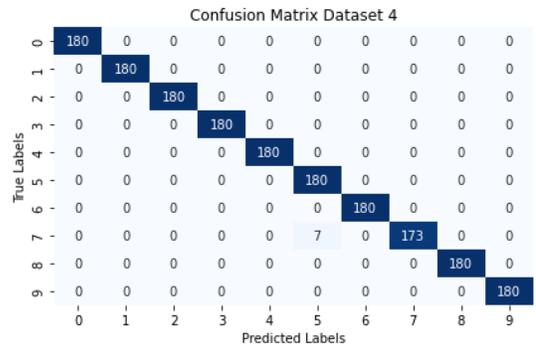

(a)
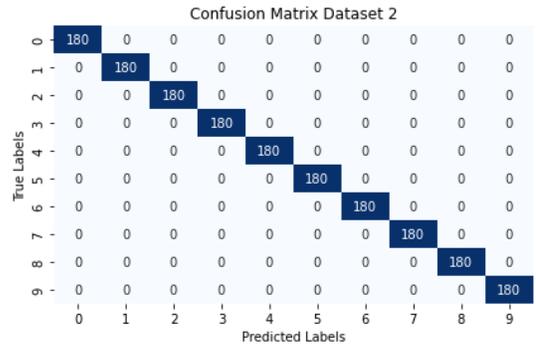
(b)

(c)

(d)

Figure 10. The confusion matrices of the four bearing datasets.

tions of different layers of the CNN architecture are further analyzed and displayed across eight subplots. In these visualizations, particularly in the convolutional layers (Layer: conv2d to conv2d3 shown in Figure 13), a remarkable separation between clusters is observed. Each cluster represents a different fault state, indicating that the convolutional layers are effectively extracting features specific to each fault type.

In the flattened layer (Layer: flatten), and the dense layers (Layer: dense to dense1), the clusters remain distinct and well-separated, reinforcing the capability of our model to differentiate between various fault states. A notable observation is in the t-SNE visualization of the final dense layer (Layer: dense2 with softmax activation for classification), where clusters appear as curved lines and dots, a result of the transformation of the feature space by the softmax layer. This distinct representation in the final layer highlights the model's ability to classify faults accurately.

3.2.3. Comparison with time-frequency analysis methods under different operating conditions

The CWT with the morlet wavelet and STFT methods are used to be compared with SSWT, which are often regarded as useful for processing transient and nonstationary signals. To verify the robustness and the generalization performance of the proposed method, tests under different operating conditions are performed. In this experiment, a dataset from one load is utilized for training, and two datasets from another load are used for testing. For instance, the notations $A \mapsto B$ and $A \mapsto C$ mean that the dataset of load A is used for training and the dataset of loads B and C are used, respectively, for testing. Ten different health conditions are present in each dataset under a single load; Table 3 has already provided the specifics. More precisely, loads 1, 2, and 3 are connected to datasets A, B, and C, respectively. The TF representations from the methods are randomly projected and trained with the selected CNN model for a fair comparison. In addition, the suggested model's, denoted SSWT-RP(Aug), effectiveness is evaluated using augmented data. The comparison seeks to demonstrate how data augmentation affects the model's overall performance, as exhibited by increased accuracy and improved generalization. The results displayed in Table 5 demonstrate the accuracies of tests across domains. Notably, compared to existing TF approaches, our suggested method with augmentation shows clear superiority. It is important to note that a significant accuracy difference still exists without using data augmentation techniques, underscoring the robust feature representation capability of SSWT-RP.

3.2.4. Comparison with other methods

The results presented in Table 6 offer compelling evidence of the superiority of our proposed algorithm compared to existing state-of-the-art methods on the same dataset. The sug-

gested method surpasses alternative methods in classification performance. It achieves classification accuracies of 100%, 100%, 99.22%, and 99.61% for datasets 1, 2, 3, and 4, respectively, resulting in an average accuracy of 99.71% across all datasets. The only misclassifications were 0.27% and 0.44% for the OR7 and B14 fault types, respectively, in the load (2 hp), with OR7 being mistaken for OR14 and B14 for B21. Similarly, in the load (3 hp), the sole misclassification was 0.38%, with B21 being misclassified as IR14.

When compared to other models, the suggested CNN method's evaluation on the CWRU bearing dataset demonstrated its outstanding generalization capability and resilience. Notably, the CNN model showed higher resistance to noise and changes in the vibration data compared to the other models, as evidenced by its much lower standard deviation in performance measures. Our algorithm achieved significantly higher accuracy rates on all four datasets, with rapid convergence after only four epochs of training, and remained stable throughout the remaining training epochs. The CNN architecture's built-in hierarchical feature learning, capable of capturing complex Time-Frequency patterns in the 2D images obtained from time-series vibration data through SSWT, was vital in enhancing the model's ability to recognize subtle fault signals.

3.3. Case 2: IMS dataset

In this section, the proposed approach is applied to another motor bearing dataset provided by the Center for Intelligent Maintenance Systems (IMS) (J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services, 2007), to further validate the method and test its effectiveness. Figure 14a and Figure 14b display the positioning of the sensors and the mechanical structure of the system, respectively (J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services, 2007).

Four bearings were arranged along a single shaft to form the bearing test rig. An AC motor, coupled to the shaft via a rubber belt, drove the shaft at a constant speed of 2000 rpm. A spring mechanism provided a radial force of approximately 26690 N on the shaft and bearing. An oil circuit system that controlled the lubricant's flow ratio and temperature was used to lubricate the bearings. The bearings were all double-row Rexnord ZA-2115 bearings. Two PCB 353B33 High Sensitivity Quartz ICP accelerometers for each bearing (x- and y-axes) were placed on the bearing housing for dataset 1, but for datasets 2 and 3, only one accelerometer for each bearing. All failures occurred after the bearing had completed more than its intended lifetime of more than 100 million revolutions.

The datasets describe tests up to failure and include files with snapshots of 1-second vibration signals taken at predetermined intervals (every 5 or 10 minutes). The NI DAQ Card 6062E was used to collect the data. These files contain 20,480 data points that were timestamped for collecting events and sampled at 20 kHz. Three tests were conducted in this exper-
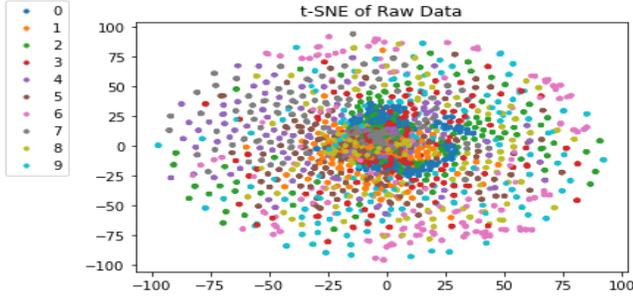
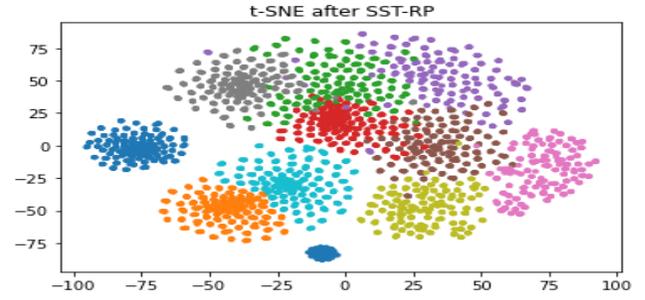Figure 11. Label colors and t-SNE of Raw Data

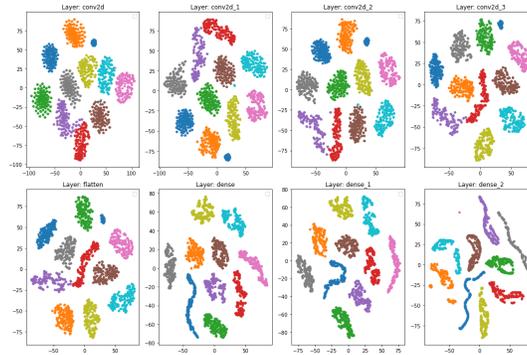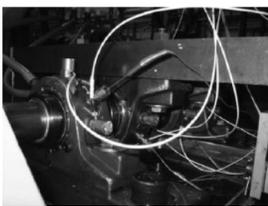Figure 12. t-SNE after SSWT and RP

Figure 13. t-SNE visualization of CNN layers

Method	$A \mapsto B$	$A \mapsto C$	$B \mapsto C$	$C \mapsto A$	Average
STFT-RP	87.5 ± 2.33	89.12 ± 1.54	81.21 ± 0.87	90 ± 0.99	86.96
CWT-RP	93.39 ± 0.25	91.33 ± 1.11	87.21 ± 0.98	89.78 ± 0.25	92.68
SSWT-RP	95.12 ± 0.96	94.87 ± 0.74	94.48 ± 0.30	95.66 ± 0.15	95.04
SSWT-RP(Aug)	$\mathbf{98.5 \pm 0.54}$	$\mathbf{99.1 \pm 0.22}$	$\mathbf{96 \pm 0.34}$	$\mathbf{97.16 \pm 0.12}$	**97.68**

Table 5. Comparison of different TF methods under noise and different loads/speeds

(a) The bearing test rig and sensor placement.
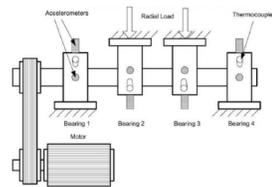
(b) The mechanical system's structure.

Figure 14. The experimental setup (J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services, 2007).

iment, resulting in a roller element defect in bearing 4 and an inner race defect in bearing 3 at the end of test 1. In the other tests, 2 and 3 respectively, bearings 1 and 3 had outer race failure. The test folders contain several files each, and only 30 out of 750 files were randomly selected for each condition type, to demonstrate the reliability and the generalization ability of the proposed method, without the need for large amounts of data thanks to the data augmentation techniques. Each revolution has 600 data points at a sampling rate of 20 kHz and a rotational speed of 2000 RPM. The size is set to be twice the rotation period, or 1200 data points. Every condition now contains 512 segments, which leads to 3072 segments after applying the four data augmentation techniques in the proposed method. The training dataset forms 62.5% of the total dataset, and the remaining 37.5% forms the testing dataset. Table 7 lists the health conditions and sample numbers for training and testing.

3.3.1. Results of Signal Transformation

The same procedure is applied to augment the vibration signals for the IMS dataset, preserving the same parameters of the techniques used above. Following the chosen signal segmentation of the 1200 data points used in the subsection above, the results of the signal-to-image conversion of the four bearing health conditions using the SSWT Transform are shown in Figure 15. The converted images have a size of 128×1200

Method	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Average
Altman,Mathew(FFT-SVM)	68.6	60	67.6	68.4	66.15
Altman,Mathew(FFT-ANN)	82.2	82.6	77	76.9	79.68
Gao(WDCNN)	99	98	99	99	98.75
Chen(CSCoh)	99.39 ± 0.20	99.33 ± 0	99.33 ± 0	97.68 ± 0.98	98.93
Yang(HSA-CNN)	99.75 ± 0.36	99.65 ± 0.33	98.85 ± 0.31	99.66 ± 0.12	99.34
Proposed Method	$\mathbf{100 \pm 0}$	$\mathbf{100 \pm 0}$	$\mathbf{99.22 \pm 0.36}$	$\mathbf{99.61 \pm 0.12}$	**99.71**

Table 6. Comparison of different methods with our proposed method

Health Conditions	Label	Training samples	Testing samples
Healthy	0	1920	1152
Inner Race	1	1920	1152
Outer Race	2	1920	1152
Roller	3	1920	1152
Total Samples	–	**7680**	**4608**

Table 7. Bearing health conditions and class labels of IMS
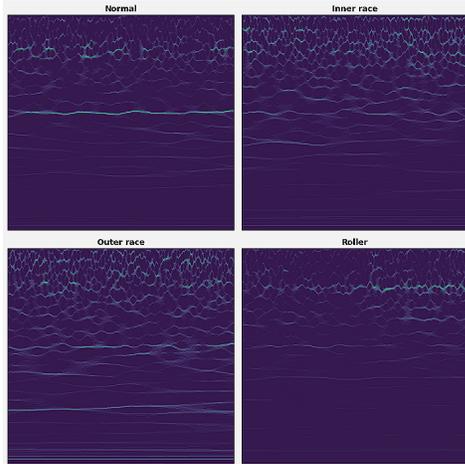

Figure 15. The SSWT of all IMS bearing health states.

pixels each.

It can be seen from the projected images that the different fault conditions appear completely distinct from one another. Random projection is the last stage of data transformation. The results of the random projection of the obtained images in Figure 15 are presented in Figure 16.

3.3.2. Analysis and Results

The IMS dataset is used to train our CNN model using the same architecture as in Case 1. The training accuracy and loss throughout 20 training epochs are shown in the line chart of Figure 17(a).

A quick learning rate was evident from the rapid drop in training losses from epochs 1 to 6. The curve began to stabilize around epoch 7, after which the model consistently attained
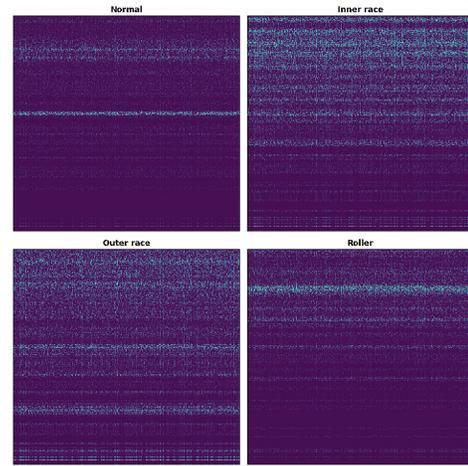

Figure 16. The RP of the obtained SSWT bearing health states of IMS.

100% accuracy. The confusion matrix in Figure 17b, on the other hand, shows the results of the testing dataset. The algorithm made no single mistake when classifying all the bearing conditions, resulting in a remarkable 100% total testing accuracy.

Significant insights are revealed by the investigation of the t-SNE clustering results from the CNN architecture used with the IMS Bearing Dataset. The raw data's t-SNE plot (Figure 18) does not show clear clusters corresponding to various fault states; instead, it appears to have class labels spread at random. However, after employing the SSWT and RP, the clusters emerged and took on more distinct patterns (Figure 19).

The convolutional layers (Layer: conv2d8-conv2d9) in Figure 20 reveal different clusters for each type of fault, demonstrating the model's capacity to identify fault-related information. The flattened layer (Layer: flatten2), where the distinct separation of fault conditions continues, reflects this. The dense layers (Layer: dense6-dense7) maintain cluster separation, affirming the architecture's effectiveness in generating refined representations that preserve fault distinctions. However, an unexpected result emerges in the t-SNE visualization of the last dense layer (Layer: dense8), which includes a softmax activation for classification. Here, two clusters (Class 0 and 2) are scattered, and two (Class 1 and 3) are repre-

(a)

(b)

Figure 17. The loss and accuracy curves (a) and the confusion matrix (b) of the IMS Dataset.

sented by a dot, potentially due to the transformative effects of the softmax layer on original feature representations and the distribution change. It is important to remember that using t-SNE on the softmax output layer may not always yield meaningful results. The combination of the t-SNE clustering results offers strong evidence of the model's ability to detect complex fault-related features. The architecture's capability to accurately reflect the various fault scenarios is demonstrated by the clusters' continuous separation across layers.

3.3.3. Comparison with Other Methods

Alternative deep learning (DL) methodologies have been selected to compare prediction accuracy in this situation and evaluate the efficacy of the proposed CNN model. The compared DL approaches include a LeNet-based CNN (L. Wen et al., 2018), a method with good anti-noise and domain adaptation ability (W. Zhang, Peng, Li, Chen, & Zhang, 2017), a multi-sensor information fusion and 2DCNN method (Wang, Wang, Wang, Li, & Song, 2021), raw sensor data using deep neural networks considering temporal coherence (R. Zhang, Peng, wu, Yao, & Guan, 2017), and an enhanced integrated filter network (Wu, Tao, Yang, Xie, & Li, 2022), along with the method proposed in this paper. The mean prediction accuracy serves as the final criterion for evaluation in the comparison. Table 8 contains the summary information of the results. The comparison demonstrates the clear advantage of the suggested method over alternative deep learning techniques. The average prediction accuracy is a remarkable 99.98%, frequently exceeding 100% in most model runs. The following approaches' prediction accuracies are: Enhanced integrated filter network (99.90%), LeNet-based CNN (99.79%), Multi-sensor information fusion and 2DCNN (99.90%), DNN considering temporal coherence (99.94%), and Anti-noise and domain adaption ability (96%). The results on the IMS dataset confirm the outstanding performance of the proposed CNN method.

4. CONCLUSION & FUTURE RESEARCH WORK

In this study, a new fault diagnosis framework called SSWT-RP-CNN is proposed to increase the precision and robustness in identifying rolling bearing faults in mechanical systems. The capacity of the CNN architecture's built-in hierarchical feature learning to capture complicated Time-Frequency patterns in the 2D images obtained from time-series vibration data through SSWT was vital in enhancing the model's ability to recognize tiny fault signals. Our method was validated on two large experimental datasets, resulting in remarkable results on both datasets, achieving a high accuracy rate. By training and testing the approach with a large amount of data, highly precise and reliable results were obtained. Our method successfully detected various types of bearing faults of different size damages, including inner race, outer race, and ball faults, with a remarkable level of accuracy. These results show that our method is a promising and dependable strategy for bearing diagnosis applications in real-world scenarios due to its ability to generalize effectively to unknown bearing conditions.

The focus on bearing fault diagnosis in this work provides a potential avenue for further study. Exploring bearing prognostics, particularly in predicting remaining useful lifetime (RUL), is a natural next step in this research. Even though the study was successful in accurately identifying the faults, concentrating on forecasting bearing lifespan can provide useful advice for operational and maintenance decisions. In the future, the focus is planned to be on estimating the RUL. This evolution offers enhanced resource planning and decision-making capabilities, by offering a better estimation of how long you can keep the component before it fails.

REFERENCES

Achlioptas, D. (2003). Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of Computer and Sys-*
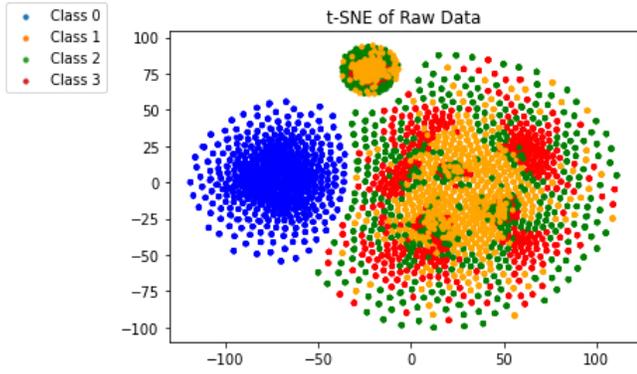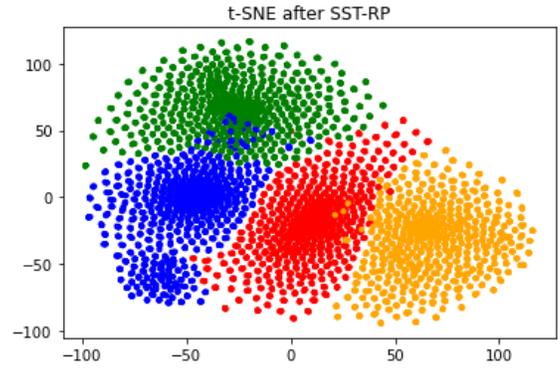
Figure 18. Label colors and t-SNE of Raw Data

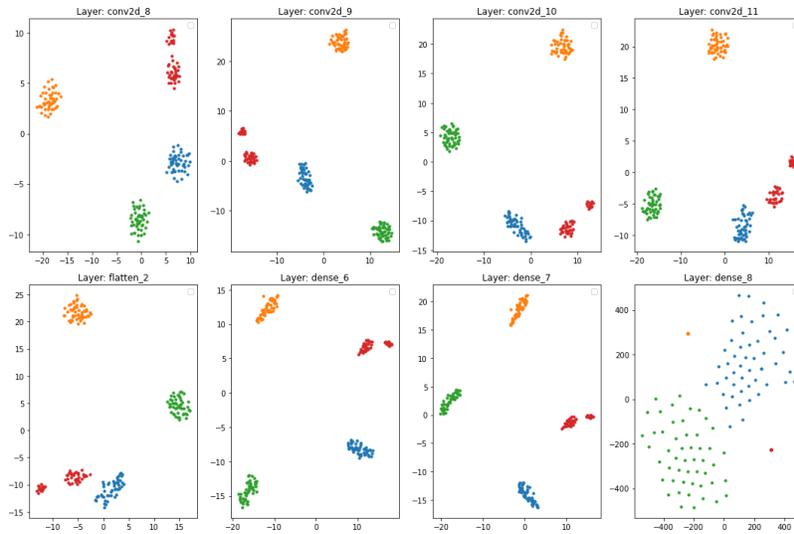Figure 19. t-SNE after SSWT and RP

Figure 20. t-SNE visualisation of CNN layers

Method	Average accuracy
Anti-noise and domain adaptation ability(W. Zhang et al., 2017)	96%
LeNet based CNN(L. Wen et al., 2018)	99.79%
Multi-sensor information fusion and 2DCNN(Wang et al., 2021)	99.90%
DNN considering temporal coherence(R. Zhang et al., 2017)	99.94%
Enhanced integrated filter network(Wu et al., 2022)	99.90%
Proposed Method	**99.98%**

Table 8. Comparison results.

tem Sciences, *66*(4), 671-687. Retrieved from `https://www.sciencedirect.com/scienc e/article/pii/S0022000003000254` (Special Issue on PODS 2001) doi: https://doi.org/10.1016/S0022-0000(03)00025-4

Altman, J., & Mathew, J. (2001). Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *15*(5), 963-977. doi: https://doi.org/10.1006/mssp.2001.1410

Auger, F., & Flandrin, P. (1995). Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, *43*(5), 1068-1089. doi: 10.1109/78.382394

Auger, F., Flandrin, P., Lin, Y.-T., McLaughlin, S., Meignen, S., Oberlin, T., & Wu, H.-T. (2013). Time-frequency reassignment and synchrosqueezing: An overview. *IEEE Signal Processing Magazine*, *30*(6), 32-41. doi: 10.1109/MSP.2013.2265316

Bengio, Y. (2016). *Deep learning*. London, England: MIT Press.

Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: Applications to image and text data. In *Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining* (p. 245–250). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/502512.502546` doi: 10.1145/502512.502546

Bouvrie, J. (2006). *Notes on convolutional neural networks.*

Chen, Z., Mauricio, A., Li, W., & Gryllias, K. (2020, 01). A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. *Mechanical Systems and Signal Processing*, *140*. doi: 10.1016/j.ymssp.2020.106683

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics. doi: 10.1137/1.9781611970104

Daubechies, I. (2000, 11). A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models.

Gai, J., Shen, J., Hu, Y., & Wang, H. (2020). An integrated method based on hybrid grey wolf optimizer improved variational mode decomposition and deep neural network for fault diagnosis of rolling bearing. *Measurement*, *162*, 107901. doi: https://doi.org/10.1016/j.measurement.2020.107901

Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2020, 01). Fundamental concepts of convolutional neural network. In (p. 519-567). doi: 10.1007/978-3-030-32644-936

Hu, Q. (2006). Intelligent diagnosis for incipient fault based on lifting wavelet package transform and support vec-

tor machines ensemble. *Chinese Journal of Mechanical Engineering*, *42*, 16.

J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services. (2007). *Ims, university of cincinnati. "bearing data set", nasa prognostics data repository, nasa ames research center, moffett field, ca.* `https://www.nasa.gov/content/prognost ics-center-of-excellence-data-set-rep ository/Bearings.`

Johnson, W., & Lindenstrauss, J. (1984, 01). Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics*, *26*, 189-206. doi: 10.1090/conm/026/737400

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization.* arXiv. Retrieved from `https://arxiv.org/abs/1412.6980` doi: 10.48550/ARXIV.1412.6980

Kononenko, I., & Kukar, M. (2007). Chapter 7 - data preprocessing. In I. Kononenko & M. Kukar (Eds.), *Machine learning and data mining* (p. 181-211). Woodhead Publishing. doi: https://doi.org/10.1533/9780857099440.181

LeCun, Y., Bengio, Y., & Hinton, G. (2015, May). Deep learning. *Nature*, *521*(7553), 436–444. Retrieved from `http://dx.doi.org/10.1038/nature14539` doi: 10.1038/nature14539

Lee, I., Chen, D., Wu, Y., & Jamerson, C. (1989). Modeling of control loop behavior of magamp post regulators. In *Conference proceedings., eleventh international telecommunications energy conference* (p. 20.1/1-20.1/8 vol.2). doi: 10.1109/INTLEC.1989.88338

Lei, Y. (2008, 01). Fault diagnosis based on novel hybrid intelligent model. *Chinese Journal of Mechanical Engineering - CHIN J MECH ENG*, *44*. doi: 10.3901/JME.2008.07.112

Liang, X., Zuo, M. J., & Feng, Z. (2018). Dynamic modeling of gearbox faults: A review. *Mechanical Systems and Signal Processing*, *98*, 852-876. doi: https://doi.org/10.1016/j.ymssp.2017.05.024

Liu, Y., Li, K., & Chen, P. (2018, 03). Fault diagnosis for rolling bearings based on synchrosqueezing wavelet transform. *Zhongguo Jixie Gongcheng/China Mechanical Engineering*, *29*, 585-590. doi: 10.3969/j.issn.1004-132X.2018.05.013

M.Benbrahim. (2014). *Classification automatique des signaux sismiques*. Omniscriptum.

Pham, M.-T., Kim, J.-M., & Kim, C.-H. (2021a). 2d cnn-based multi-output diagnosis for compound bearing faults under variable rotational speeds. *Machines*, *9*(9). doi: 10.3390/machines9090199

Pham, M. T., Kim, J.-M., & Kim, C. H. (2021b). Efficient fault diagnosis of rolling bearings using neural network architecture search and sharing weights. *IEEE Access*, *9*, 98800-98811. doi: 10.1109/AC-

CESS.2021.3096036

Renwick, J. T., & Babson, P. E. (1985). Vibration analysis—a proven technique as a predictive maintenance tool. *IEEE Transactions on Industry Applications*, *IA-21*(2), 324-332. doi: 10.1109/TIA.1985.349652

Ricci, R., & Pennacchi, P. (2011). Diagnostics of gear faults based on emd and automatic selection of intrinsic mode functions. *Mechanical Systems and Signal Processing*, *25*(3), 821-838. Retrieved from https://www.sciencedirect.com/science/article/pii/S0888327010003201 doi: https://doi.org/10.1016/j.ymssp.2010.10.002

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 60. doi: 10.1186/s40537-019-0197-0

Smith, W. A., & Randall, R. B. (2015). Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, *64-65*, 100-131. Retrieved from https://www.sciencedirect.com/science/article/pii/S0888327015002034 doi: https://doi.org/10.1016/j.ymssp.2015.04.021

Wang, J., Wang, D., Wang, S., Li, W., & Song, K. (2021). Fault diagnosis of bearings based on multi-sensor information fusion and 2d convolutional neural network. *IEEE Access*, *9*, 23717-23725. doi: 10.1109/ACCESS.2021.3056767

Wang, J., Zhuang, J., Duan, L., & Cheng, W. (2016). A multi-scale convolution neural network for featureless fault diagnosis. In *2016 international symposium on flexible automation (isfa)* (p. 65-70). doi: 10.1109/ISFA.2016.7790137

Wen, J., Gao, H., Li, S., Zhang, L., He, X., & Liu, W. (2015, 10). Fault diagnosis of ball bearings using synchrosqueezed wavelet transforms and svm. In (p. 1-6). doi: 10.1109/PHM.2015.7380084

Wen, L., Li, X., Gao, L., & Zhang, Y. (2018). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, *65*(7), 5990-5998. doi: 10.1109/TIE.2017.2774777

Wu, K., Tao, J., Yang, D., Xie, H., & Li, Z. (2022, 06). A rolling bearing fault diagnosis method based on enhanced integrated filter network. *Machines*, *10*, 481. doi: 10.3390/machines10060481

Zan, T., Wang, H., Wang, M., Liu, Z., & Gao, X. (2019). Application of multi-dimension input convolutional neural network in fault diagnosis of rolling bearings. *Applied Sciences*, *9*(13). doi: 10.3390/app9132690

Zhang, R., Peng, Z., wu, L., Yao, B., & Guan, Y. (2017, 03). Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence. *Sensors*, *17*, 549. doi: 10.3390/s17030549

Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, *17*(2). Retrieved from https://www.mdpi.com/1424-8220/17/2/425 doi: 10.3390/s17020425

BIOGRAPHIES

Boubker Najdi's academic journey began in 2018 when he earned a B.Eng. degree in Mechatronics and Embedded Systems from the Faculty of Sciences, Fez. He continued his educational pursuits by obtaining an M.Eng. in Smart Industry from the University of Sidi Mohamed Ben Abdellah, Fez, Morocco, in 2020.

Currently, he is a Data Scientist and a Ph.D. student at the Faculty of Sciences, Fez. His areas of research expertise encompass a wide range of topics, including robotics, artificial intelligence, and predictive maintenance.

For inquiries or contact, you can reach him at the following email address: aboubakr.najdi@usmba.ac.ma.

Mohammed Benbrahim received the B.Eng. degree in electromechanical engineering from the Higher National School of Mines in 1997 and the M.Sc. and Ph.D. degrees in automatic and industrial informatics from the Mohammadia School of Engineers in 2000 and 2007, respectively.

Currently, he is a Full Professor at the Department of Physics and the program coordinator for the Master's program in Smart Industry at the Faculty of Sciences, Sidi Mohamed Ben Abdellah University. His research interests include robotics, automatic control, intelligent systems, predictive maintenance, modeling, and optimization. He can be contacted at email: mohammed.benbrahim@usmba.ac.ma.

Mohammed Nabil Kabbaj is a Full Professor at the Faculty of Sciences, University of Fez, where he is the Program Coordinator of the Mechatronics and Embedded Systems Bachelor's program. His research interests include control engineering, fault detection, and diagnosis of complex systems. Before joining the University of Fez, he received a Ph.D. degree from the University of Perpignan in 2004 and was a postdoctoral researcher at LAAS-CNRS in Toulouse. He can be contacted at email: n.kabbaj@usmba.ac.ma.